Commentary

Methodological Biases in Estimating the Burden of Out-of-Pocket Expenses

Dana P. Goldman and James P. Smith

In a recent issue of *Health Services Research*, Gross, Alecxih, Gibson, et al. (1999) (hereafter referred to as Gross et al.) argued that out-of-pocket health care expenses impose a substantial burden on the elderly, especially the poor elderly. They find that the elderly spend 19 percent of their income on medical care and, more importantly, that the average share rises to 35 percent for the poor elderly. Their estimates are based on simulations from the 1993 Medicare Current Beneficiary Survey (MCBS) but are similar to estimates by Moon, Kuntz, and Pounder (1996) from the 1987 National Medical Expenditure Survey using the same methodology. Gross et al. conclude from these results that Congress should increase the financial protection against high out-of-pocket expenses for the elderly.

Unfortunately, the methodology used by Gross et al. substantially overstates the true burden of out-of-pocket health care expenses for the elderly, and this overstatement is particularly large for the poor elderly. Their approach also distorts the real policy issues that exist in providing elderly Americans adequate protection against the possibility of significant medical expenses during their old age. In this commentary, we describe the source and degree of the biases that exist in their research methodology.

ESTIMATING THE BURDEN OF OUT-OF-POCKET EXPENSES

Gross et al. measure the burden of health care expenses as an average cost share—that is, the average across beneficiaries of the ratio of out-of-pocket spending to income. If reported income is R_i and reported out-of-pocket spending is OOP_I, the share is

This research was sponsored by the National Institute on Aging (P01AG08291).

Dana P. Goldman, Ph.D. is Senior Economist, RAND, 1700 Main Street, Santa Monica, CA 90407-2138. James P. Smith, Ph.D. is Senior Economist, RAND, Santa Monica, CA.

$$share = \frac{1}{N}\sum_i \frac{OOP_i}{R_i}, \tag{1}$$

where N is the number of elderly beneficiaries in the entire MCBS or some subgroup, such as those living in poverty. An alternative measure of burden—one rejected by Gross et al.—computes burden as the (within-group) ratio of average expenditures to average income, i.e.,

$$\widetilde{share} = \frac{\frac{1}{N}\sum_i OOP_i}{\frac{1}{N}\sum_i R_i}. \tag{2}$$

Using the ratio of two highly skewed distributions to characterize the burden of many elderly is always problematic. This problem is exacerbated in the Gross et al. method when bias and misreporting exist in survey estimates of income—the denominator in their cost burden estimate. This bias either does not exist or is much smaller when the alternative cost burden measure in Equation 2 is used instead.

Their share estimate is greatly affected by relatively few outliers that largely reflect measurement error. To demonstrate this point, we reproduce the results from Gross et al. in Table 1 using the 1995 MCBS Cost and Use file. We estimate an overall mean share of 20 percent and a share for the poor of 34 percent (shown in the column labeled "Censored Share"), quite close to their estimates of 19 percent and 35 percent, respectively.

One reason for concern involves the fraction of cases where the cost share exceeds 1, implying that households spend more on medical care than they earn in income. A total of 11.5 percent of poor MCBS respondents report out-of-pocket expenses greater than their total income, and some beneficiaries

Table 1: 1995 Burden of Out-of-Pocket Expenses by Poverty Status

Poverty Status	*No. of Beneficiaries*	*Percent with OOP > Income*	*Uncensored Share*	*Censored Share*	$\widetilde{share}$
Poor	6,152,379	11.5%	60%	34%	33%
Near poor	2,970,950	1.9	26	25	25
Low income	7,623,406	0.7	22	21	21
Middle income	10,245,375	0.3	14	13	13
High income	4,837,223	0.1	7	7	6
Overall	31,829,333	2.7	25	20	13

Note: Poverty status and expenses follow definitions of Gross et al. Data are for noninstitutionalized Medicare beneficiaries 65 or older. By comparison, Gross et al. report censored shares of 35 percent (poor), 23 percent (near poor), 22 percent (low income), 17 percent (middle income), 10 percent (high income), and 19 percent (overall).

report out-of-pocket expenses more that 40 times their income. Gross et al. deal with this problem of implausible values of cost shares with an ad hoc adjustment of capping the individual cost shares at 1.0. Their reported share is actually a censored estimate,

$$\textit{censored share} = \frac{1}{N}\sum_i \min\left(\frac{OOP_i}{R_i}, 1\right). \tag{1'}$$

The effects of censoring of cost shares are shown in Table 1. The overall share falls from 25 percent to 20 percent after censoring at 1. The effect on poor beneficiaries is quite dramatic, reducing the average share from 60 percent to 34 percent, indicating how sensitive their method is to these extreme and implausible values. Since people in fact do not spend all their income on health care, these extreme values indicate that either medical expenses are overstated or income is understated for many households.

The final column of Table 1 lists burden estimates from the alternative estimator in Equation 2 obtained by dividing mean expenditures by mean income. For the *overall* sample, one advantage of $\widetilde{share}$ is that the denominator will not be biased by measurement error because errors in individual values will average out to 0 in a large data set. Thus, the difference between this alternative burden estimate of 13 percent and the Gross et al. estimate of 20 percent indexes their overstatement of overall burden. While within-income-group burden estimates are approximately the same with the two alternative estimators, we demonstrate below that both estimates are biased upward, especially for low-income households.

Bias and Misreporting of Income

The quality of estimates of cost shares depends both on the quality of measurement of medical expenditures and income. The MCBS, which links Medicare claims to survey-reported events, provides an extremely comprehensive profile of out-of-pocket health care expenses. Unfortunately, MCBS income data are not nearly as complete, as it asks only a single question about total income for the respondent and spouse, if married. It misses income from other people, such as children, who may be living in the household and who can help pay some expenses. Second, combining all income into one question results in both a significant bias (typically underreporting) and misreporting (or random measurement error).

All surveys suffer from some problems of measurement error in income. The Current Population Surveys (CPS) are the most widely used source to monitor income changes by year in the United States and represent a useful standard of comparison to health surveys such as the MCBS. The income data in the CPS are only 90 percent of benchmark estimates obtained from reliable external controls, indicating that even a household survey that concentrates on income measurement underreports income by an average of 10 percent (U. S. Bureau of the Census 1998). Moreover, we calculate that MCBS income is 10 percent lower than CPS family income and 31 percent lower than CPS household income, so that on average MCBS understates family income by about 20 percent and household income by 41 percent.[1]

Even when income is reported without bias, the problem of random measurement error remains. One way to gauge its magnitude is to use repeated reports of income sources that do not change or change in a predicable way. Since the "true" income growth is known, social security may represent the ideal income source to gauge respondents' ability to report income accurately. By age 70, social security income is fixed legislatively and, without changes in family composition due to divorce, separation, or death, is only revised across calendar years by a universal cost of living adjustment (COLA).

To measure income misreporting, we matched respondents across two successive CPS March panels for years 1996 and 1997.[2] Given our CPS sample restrictions, social security income should only change due to a COLA. If respondent reports were completely accurate, the COLA-adjusted wave one and actual CPS wave two social security income reports would be identical. Differences between them reflect only measurement error.

Table 2 lists percentiles of social security income when households are either ranked by 1996 or 1997 social security income. When ranked by their 1996 income, the mean income of the bottom 10 percent is $2,874 in 1996. Even though only COLA adjustments occurred, these same households report a mean income of $4,371 in 1997. Ranking by 1996 income understated

Table 2: Measurement Error in Reported Social Security Income in the CPS

	Percentiles Based on 1996 Income		*Percentiles Based on 1997 Income*	
Income Percentile	*Mean Income, 1996*	*Mean Income, 1997*	*Mean Income, 1996*	*Mean Income, 1997*
10% or less	$2,874	$4,371	$4,554	$2,732
25% or less	3,985	5,284	5,266	3,947
90% or more	16,615	13,544	12,854	17,576
Mean	8,139	8,441	8,139	8,441

their 1996 income by 40 percent. A parallel pattern appears when we rank these same households instead by their 1997 income. Now, the bottom 10 percent (approximately the poverty population) has a 1997 mean income of $2,732 but an income of $4,554 in 1996.

With measurement error, ranking households by incomes in any year results in an understatement of the true income of households ranked at the bottom. But this ranking is precisely what Gross et al. do in computing cost shares for the poverty population in a survey with more income measurement error than the CPS. Thus, their estimates of cost burden for the poverty population are overstated by both understatement of income (bias) and random income misreporting.

BIAS IN THE OVERALL BURDEN DUE TO MISREPORTING

To measure the bias in their censored share estimate, we conducted a straightforward Monte Carlo simulation. Reported income (R_i) is assumed to be a multiplicative function of true income (T_i) and measurement error (E_i):

$$R_i = b \cdot T_i \cdot E_i. \tag{3}$$

Two measurement problems bias estimates of cost burden—bias and random measurement error. The expected value of reported income is bT_I,[3] so b measures the systematic understatement of income. The degree of random measurement error is specified using the R^2 from a (linear) regression of R_i on T_i. If R^2 is close to 1, no measurement error exists, while an R^2 near 0 implies only noise in the data. In simulations, we assume that all people spend a known and fixed share ($h = .10$) of their true income on health care. However, due to bias and measurement error in income, the observed share will vary across the sample.[4] The degree to which the estimates deviate from 10 percent measures the bias in the Gross et al. methodology.

Table 3 lists simulation results obtained for a range of plausible values of b and R^2.[5] When $b = 1$ and $R^2 = 1$, there is no underreporting or measurement error in reported income, and all estimated shares are equal to their true value of 10 percent. However, the presence of even moderate measurement error—even with no systematic underreporting ($b = 1$)—can induce a significant overstatement of the overall burden. For example, with $b = 1$, measurement error of 0.6 produced a censored share estimate of 20 percent. In contrast, the alternative estimator equation, Equation 2, is not

Table 3: Bias in Burden Estimates for Various Levels of Income Misreporting

Systematic Underreporting (b)	*Measurement Error* (R^2)	*True Share* (h)	*Censored Share*	$\widetilde{share}$
0.6	0.6	10.0%	36.3%	16.7%
	0.8	10.0	20.6	16.7
	1.0	10.0	16.7	16.7
0.8	0.6	10.0	23.6	12.5
	0.8	10.0	15.5	12.5
	1.0	10.0	12.5	12.5
1.0	0.6	10.0	19.9	10.0
	0.8	10.0	12.4	10.0
	1.0	10.0	10.0	10.0

subject to this bias due to measurement error alone, as can be seen by the consistent estimates given by $\widetilde{share}$ when $b = 1$, regardless of the level of measurement error.

The remaining rows of Table 3 demonstrate that the degree of bias is affected by both underreporting and measurement error. For example, even with no measurement error ($R^2 = 1$), a value of $b = 0.8$ (a 20 percent understatement of income) would produce an estimate of the censored share of 12.5 percent. Since understatement of income is larger in the MCBS, the Gross et al. estimates are also biased by this factor. Finally, while the alternative estimator $\widetilde{share}$ does little to mitigate the impact of systematic underreporting, it yields a less biased estimate of the censored share in the face of random measurement error.

MISREPORTING AS AN EXPLANATION FOR DIFFERENCES BY INCOME

Gross et al. also argue that the burden is much higher for low-income beneficiaries. Unfortunately, their estimates of burden by income class are even more biased by income misreporting. Essentially, when we rank people by their income, those ranked at the very bottom must necessarily have higher amounts of systematic underreporting of income and measurement error. As an example, Table 4 shows the overall burden estimates for subgroups of the population for $b = 1.0$ and $R^2 = 0.70$. By choosing $b = 1$, we are assuming no systematic bias, but we do allow for a moderate amount of measurement error in income. Even though the true burden is the same for all groups, the censored estimate induces a gradient that suggests a much higher burden for the low-income elderly.

How large is the bias in the Gross et al. estimates of cost burden? The true overall share can be obtained from the ratio of mean expenditures to mean income after we adjust MCBS mean income by its average income understatement. Table 1 indicates that the aggregate ratio of MCBS expenditures to income is .13, while mean income understatement ranges from 20 percent for family income to 41 percent for household income. These parameters imply that the true overall share lies between 8 percent and 10 percent, less than half of the 19 percent overall cost burden reported by Gross et al. In addition to the impact of average income understatement, the overstatement of cost burden for the poverty population depends also on the amount of random measurement error in the data. Based on a comparison of social security income in the CPS, we estimate this bias in the range of 40 percent so that our estimate of cost burden for the poor ranges from 11 percent to 16 percent, once again less than half that reported by Gross et al.

CONCLUSION

In this commentary, we demonstrate that the Gross et al. approach results in a significant overstatement of the cost burden of medical expenses. The magnitude of this bias is particularly large for poorer households, the legitimate focus of policy concern. There is another key lesson for the health services research community. The MCBS goes to great lengths to collect accurate, high-quality measures of health services utilization and expenditures, but measures income with a single question. For many questions, the quality of measurement of socioeconomic status is as important as high quality measurement of health status, health costs, and utilization. The MCBS is an important source of

Table 4: Simulated Bias in Burden Estimates by Income Group

Reported Income Quintile	*True Burden* (h)	*Censored Burden (Eq. 1′)*
Poorest	10.0%	35%
Second	10.0%	16%
Third	10.0%	12%
Fourth	10.0%	10%
Highest	10.0%	8%

information about outcomes for the elderly and would benefit greatly from expanded socioeconomic modules. Until then, researchers must use caution in how they use and report such data.

NOTES

1. Mean income in 1995 for elderly respondents and their spouses is \$23,313 in the MCBS and \$25,854 in the CPS. Mean household income in the CPS is \$31,760 for elderly respondents.
2. To eliminate demographic reasons for changes in social security income, we restricted our sample to households where respondents were at least 70 years old in the first CPS wave and where no marital status changes or deaths occurred between the first and second CPS wave. We also required respondents to have received some social security income in each wave so that there is no ambiguity that we are dealing with program beneficiaries. Finally, cases were deleted when income was imputed in either panel wave.
3. We assume measurement error is a linear transformation of a beta-distributed random variable because this generates error with the desirable properties that it has a mean of 1 and is symmetric. Simulations using other distributions such as the normal or log normal produced the same results.
4. Because MCBS measures expenses in more detail, we assume individuals report true out-of-pocket expenses (OOP) but noisy and biased measure of income (R). Since expenses are in the numerator of the burden estimate and we do not stratify by expense, the impact of measurement error in expenses on cost burden is small. Monte Carlo simulations allowing for measurement error in expenses confirmed this.
5. For each value of b and R^2, we drew 10,000 observations for true income (T) from a log normal distribution. We then drew the measurement error (E) from a linear transformation of a beta distribution with a mean of 1 and standard error calculated to yield a predetermined R^2 for a regression of R on T. We calculated reported income (R) as in Equation 3 and the share for each observation using $OOP_i/\underset{\sim}{R_i}$, where expenses were a fixed percent (10 percent) of T. Censored shares and $\widetilde{share}$ were then calculated using Equations 1′ and 2, respectively. These simulations were repeated 100 times for each set of values for b and R^2, and the results shown in Tables 3 and 4 are averages across the 100 simulations.

REFERENCES

Gross, D. J., L. Alecxih, M. J. Gibson, J. Corea, C. Caplan, and N. Brangan. 1999. "Out-of-Pocket Health Spending by Poor and Near-Poor Elderly Medicare Beneficiaries." *Health Services Research* 34 (1, Part II): 241–54.

Moon, M., C. Kuntz, and L. Pounder. 1996. *Protecting Low-Income Medicare Beneficiaries.* Washington, DC: The Commonwealth Fund.

U. S. Bureau of the Census. 1998. *Money Income in the United States.* Current Population Reports P60-200, 1997 (with separate data on valuation of noncash benefits). Washington, DC: U. S. Government Printing Office.

Reply

Lisa Alecxih, John Corea, David Gross, Craig Caplan, Normandy Brangan, and Mary Jo Gibson

In this issue of *Health Services Research* (*HSR*), Dr. Goldman and Dr. Smith (Goldman and Smith 2001) provide a commentary to Gross, Alecxih, Gibson, et al.'s (1999) article, "Out-of-Pocket Health Spending by Poor and Near-Poor Elderly Medicare Beneficiaries." Goldman and Smith argue that our methodology of computing shares of income spent on health care significantly overstates the financial burden of out-of-pocket expenses, especially for low-income beneficiaries, because the income data used suffer from measurement errors (bias and misreporting) and because we used an inappropriate method for calculating burden. In this response, we address these criticisms and show that their proposed alternative measure systematically understates the burden of health care spending, and, in our view, is not a better measure of out-of-pocket burden.

INCOME DATA SHORTCOMINGS

Goldman and Smith perceive potential problems with income reporting in the Medicare Current Beneficiary Survey (MCBS) because 11.5 percent of poor respondents report out-of-pocket expenses greater than their income. As the authors assume that people do not spend all of their income on health care, they contend that income must be understated (or medical expenses overstated) for many of these low-income households. While we agree with Goldman and Smith that the MCBS income data may have shortcomings, we disagree with their assumption that these outliers *necessarily* imply errors in reporting income. There are several reasons why out-of-pocket expenditures could exceed reported income. For example, such spending could be the result of any of the following factors: Medicare beneficiaries using assets to